Cell Load-Aware Energy Saving Management in Self-Organizing Networks

Henrik Klessig, Albrecht Fehske, and Gerhard Fettweis
Vodafone Chair Mobile Communications Systems
Dresden University of Technology, Germany
Email: [henrik.klessig, albrecht.fehske, fettweis]@ifn.et.tu-dresden.de

Jens Voigt
Actix GmbH
Dresden, Germany
Email: jens.voigt@actix.com

Abstract—Self-organizing networks are considered to be the next generation technology for network management and, therefore, also treated as one possible way to tackle economical and ecological challenges in the future. In this paper, we extend our framework for integrated self-organizing networks by the energy saving management self-organizing network use case. In contrast to coordination among multiple use cases, this approach targets at an integrated solution, where the coordination is inherent in the optimization enabling managing multiple use cases concurrently. Moreover, it considers dynamic traffic and interference situations, and predicts cell loads for various network configurations based on measurements of receive powers and traffic conditions. Using the cell load as a proxy for evaluating network performance and base station energy consumption, we show that a joint optimization of cell individual offsets and antenna downtilts is superior to the adjustment of cell individual offsets only. Throughout a day, we observe remarkable improvements in user throughputs and reduction in network energy consumption further enhanced by switching off base stations.

Index Terms—self-organizing network; energy saving management; base station switching; LTE; cell load modeling; traffic modeling; interference modeling

I. INTRODUCTION

The ICT sector and, therein, mobile cellular networks are undoubtedly great contributors to economical and ecological issues [1], [2]. Self-organizing networks (SON) are considered to be the next generation technology for network management, i.e., network planning, optimization, and healing, and, therefore, also treated as one possible way to tackle the economical and ecological challenges in the future. The work by the Next Generation Mobile Networks (NGMN) alliance [3] and by the European 7th Framework research project Socrates [4] constitutes pioneering and remarkable achievements in the field of SON, which are mainly being standardized by that time by the Third Generation Partnership Programme (3GPP), see, e.g., [5]. The SON community is still keeping the momentum of research going within the European 7th Framework research project Semafour [6], extending the achievements by its predecessor, the Socrates project, by research into integrated SON for multi-RAT and multi-layer cellular networks.

A. State-of-the-Art SON

Amongst others, one outcome of the Socrates project is a holistic framework to discover SON use cases and possible interdependencies among network control parameters and SON algorithm implementations. On top of this framework, recommendations are made to coordinate multiple algorithms and use cases in order to prevent suboptimal or even conflicting control parameter changes leading to worse network performance [7], [8]. E.g., a policy-based approach for a tailing coordination of the mobility load balancing (MLB) and mobility robustness optimization (MRO) use cases was presented in [9]. It is agreed upon the fact that appropriate coordination mechanisms are crucial in SON design.

Conceptually, there exist three ways of SON implementation: Centralized, distributed, and hybrid. Whereas the centralized approach requires full information collected and processed by a central node, SON algorithms implemented in the individual base stations (BS) determine control parameter changes only depending on local information in the distributed approach. The hybrid SON unifies both, the centralized and distributed concepts [10].

B. Integrated SON Approach Proposed

In this paper, we extend our work on the capacity and coverage optimization (CCO), cell outage compensation (COC), and mobility load balancing (MLB) use cases in [11], [12] by the energy saving management (ESM) use case further developing our framework towards a holistic approach for integrated SON.

In this regard, we stick to the centralized SON solution, assuming knowledge about receive powers and traffic distributions to be available at a central node. This information can be obtained by measurements, e.g., performance management (PM) measurement counters of the BSs or measurements by the user equipment (UE), so-called call traces (CT), see [13], [14] for further details. We believe, that centralized algorithms operating on different time scales will be applicable for network optimization in the future, since the wireless community is currently expending efforts in research into cloud radio access networks (C-RAN) and fiber backhaul, thus enabling fast connections to data and computation centers.

In contrast to coordination among multiple SON use cases, the approach presented in this paper targets at an integrated SON: The coordination is inherent in the optimization leading to a joint optimization of multiple network control parameters while managing the aforementioned use cases concurrently. Moreover, it considers dynamic traffic and interference situations on the basis of a cell load model independently proposed.
in [15] and [16], which already has been applied to SON algorithms in, e.g., [17], [18]. For an exemplary study of the model’s accuracy, we refer to [12]. In order to optimize the mobile network in terms of energy consumption and efficiency, macro base stations are switched off. Subsequently, antenna tilt settings and user association mechanisms, i.e., initial cell selection, cell re-selection, and handover, are adjusted to compensate for possible performance degradations, as envisaged by the mechanisms ES (de)activation and ES compensation being standardized by 3GPP [19].

The paper is organized as follows: Section II introduces the system model used throughout the paper, which is considered as the basis for the ESM SON algorithm presented in Section III. Simulation results are presented in Section IV and conclusions are given in Section V.

II. SYSTEM MODEL

The system model including the definition of the cell load has already been introduced in detail in [12], and is recapitulated here briefly.

A. Network Layout and Traffic

We consider the downlink of a cellular network with \( N \) macro BSs deployed in a compact region \( R \subseteq \mathbb{R}^2 \). Users are spatially distributed according to some normalized distribution \( \delta(\cdot) \) with \( \int_R \delta(u) \, du = 1 \). Network traffic is modeled on flow level where flows represent individual data transfers of, e.g., web pages, video, audio, or general data files. We further model the arrival of flow requests as a Poisson process with intensity \( \lambda \) and flow sizes as exponentially distributed with common mean \( \Omega \). The terms \( \lambda, \Omega, \) and \( \delta(u) \) yield the mean traffic intensity distribution \( \sigma(u) := \lambda \delta(u) \) in Mbps per km².

In addition, we define network coverage based on the reference signal received power (RSRP) experienced by the UEs. Let \( p_i(u) \) denote the power received from BS \( i \) at location \( u \). We consider the corresponding coverage region

\[
L := \{ u \in R | \exists i : p_i(u) \geq p_{\text{min}} \},
\]

where \( p_{\text{min}} \) models the minimum receive power required to connect to the network. The degree of network coverage \( C \) is then defined as the fraction of users covered, i.e.,

\[
C := \int_L \delta(u) \, du.
\]

We denote the serving area or cell of BS \( i \) by \( L_i \subset L \) and further collect the cells \( L_i \) in the partition \( \mathcal{P} := \{ L_1, \ldots, L_N \} \).

B. Radio Link

1) Receiving Conditions: We assume fast and slow fading effects to be contained as averages in the location-dependent but otherwise constant functions \( p_i(\cdot) \). Similar to slow fading effects, antenna tilt angles may have a strong effect on the propagation conditions and thus on the receive powers \( p_i \). Throughout the paper, we denote the tilt angle of all antennas at BS \( i \) by \( e_i \) and collect all tilts in the vector \( e = (e_1, \ldots, e_N)^T \). To keep notation simple, we omit the dependency of the receive powers, and all corresponding terms on the tilts, i.e., we write \( p_i(\cdot) \) instead \( p_i(\cdot, e_i) \).

2) Radio Link Quality: We define the SINR \( \gamma_i \) experienced by a data flow at location \( u \) with respect to BS \( i \) as

\[
\gamma_i(u, \eta) := \begin{cases} 
\frac{p_i(u)}{\sum_{j \neq i} \eta_j p_j(u) + \theta} & \text{if } p_i(u) \geq p_{\text{min}}, \\
0 & \text{otherwise},
\end{cases}
\]

where \( \theta \) and \( p_{\text{min}} \) denote the noise power and the minimum signal power required to connect to the network, respectively. The terms \( \eta_j \in [0, 1] \) denote the loads of the interfering BSs \( j \neq i \). The corresponding achievable data rate is modeled based on the Shannon capacity, i.e.,

\[
c_i(u, \eta) = \min \left\{ a \log_2 \left( 1 + b \gamma_i(u, \eta) \right), c_{\text{max}} \right\},
\]

where the term \( c_{\text{max}} \) denotes the maximum bit rate achievable given by the highest modulation and coding scheme of the system at hand. According to [20], we incorporate an average packet scheduling gain via the parameters \( a \) and \( b \), and choose corresponding values for more spectrally efficient scheduling mechanisms, MIMO techniques, or system specific overheads.

Handover events are commonly triggered by user mobility and the slow fading process. Since both happen on a much larger time scale than a typical flow duration, we assume that a flow remains connected to a single serving BS during its entire lifetime and omit modeling handover processes.

3) Effects of Inter-Cell Interference: The radio link quality is further governed by the collection of BSs that are interfering at any given point in time. As opposed to slow and fast fading effects, these interference scenarios evolve on the same time scale as the flow dynamics and, as a result, data rates as well as cell loads of all BSs are strongly interconnected.

Instead of assuming some fixed minimal or maximal interference levels by or data rates in the neighboring cells as done in [21], we consider mutual interference among all the BSs, leading to coupled cell loads and interference levels, and, therefore to coupled data rates in the entire network. In order to capture the effect of dynamic interference scenarios on data rates, we utilize the technique proposed independently in [15] and [16], where we consider the bitrate as if flows are exposed to average interference conditions. According to the underlying M/M/1 PS queueing model, the load \( \eta_i \) corresponds to the probability that BS \( i \) is transmitting. Consequently, the term \( \sum_{j \neq i} \eta_j p_j(u) \) denotes the time averaged interference power. Note that other scheduling policies than processor sharing (as the limiting case of a Round Robin scheme) would yield other types of queueing systems; however, for mathematical tractability and simplicity, we incorporate corresponding gains via the factors \( a \) and \( b \) as explained above.

C. Resource Utilization

Considering the load density with respect to BS \( i \) to be defined as \( \kappa_i(u, \eta) := \frac{c_i(u, \eta)}{c_{\text{max}}} \), we define the average resource utilization of BS \( i \), i.e., its load, as the integral of the load
density over the serving cell area as
\[ f_i(\eta) := \min \left\{ \int_{\mathcal{L}_i} \kappa_i(u, \eta) \, du, \ 1 - \epsilon \right\}, \tag{5a} \]
with an arbitrarily small \( \epsilon > 0 \). The bound \( 1 - \epsilon \) becomes necessary when we introduce load-dependent partitions \( \mathcal{P} \) and cell areas \( \mathcal{L}_i \) later. Observe, that Eq. (5a) only gives an implicit formulation of the cell load since the achievable rates \( c_i(u, \eta) \) also depend on the load vector. Let \( f = (f_1, \ldots, f_N)^T \) denote the vector valued function with components \( f_i \). The load vector of interest is then given as solution to the system
\[ \eta = f(\eta). \tag{5b} \]

In the following, we interpret the cell partition \( \mathcal{P} \) and the antenna tilts \( \epsilon \) as free variables. Hence, the load situation in the network is given as the solution to the system \( \eta = f(\eta, \mathcal{P}, \epsilon) \).

In particular, for all tuples \( (\mathcal{P}, \epsilon) \), the System (5) always has a unique fixed point in \([0, 1]^N\) and, thus the load is well defined (refer to [15] Theorem 1). Since the solution is based on time average receive powers and spatial traffic distributions, we interpret the cell loads as time averages as well. For the network optimization approach outlined subsequently, we use the cell load computation as a tool to estimate the average BS utilizations observable in time intervals of several minutes up to a few hours. Specifically, we do not measure the loads in the individual BSs. This allows for, e.g., predicting the utilizations of the residual BSs prior to switching off.

III. ENERGY SAVING MANAGEMENT

It is well known that the cell load serves as an indicator for network performance, e.g., in terms of throughputs, congestion probability, and delays (see, e.g., [22] for details), and for the energy consumption of base stations [23], as well. Hence, we exploit the cell load as a proxy for network performance evaluation, specifically for the ESM algorithms proposed below.

A. Problem Formulation

Similar to [12] and [24], we optimize the tilts \( \epsilon \) and partition \( \mathcal{P} \) by minimizing the \( \alpha \)-parametrized objective function \( \Phi_\alpha \), i.e.,
\[ \min_{\mathcal{P}, \epsilon} \Phi_\alpha(\eta) = \begin{cases} \sum_{i=1}^N \left( \frac{(1 - \eta_i)^{1 - \alpha} - (1 - \alpha) \sum_{i=1}^N \log(1 - \eta_i)}{\alpha - 1} \right) & \text{for } \alpha \neq 1, \\ \sum_{i=1}^N \log(1 - \eta_i) & \text{for } \alpha = 1. \end{cases} \tag{6} \]
subject to \( \eta = f(\eta, \mathcal{P}, \epsilon), \ C \geq C_{\min}, \)
where we explicitly require the loads \( \eta \) to be obtained as solution to the System (5) and the RSRP coverage to be larger than a given threshold \( C_{\min} \).

In order to obtain the optimal partition \( \mathcal{P}^* \), we formally define the user association rule \( s(u, \eta) \) forming the cells
\[ \mathcal{L}_i = \left\{ u \in \mathcal{L} \mid i = s(u, \eta) \right\}. \tag{7} \]

B. Solution Proposed

We propose the SON algorithm to optimize the network regarding two different strategies resp. objectives:

1) Energy-efficiency: Given a fixed set of active BSs, for \( \alpha = 0 \), we minimize the sum of their loads. This results in a minimization of network energy consumption, if the BSs are of the same type, while still serving the traffic demand. The linear BS power model used is given in Table II. As shown in [24], the optimal partition \( \mathcal{P}^* \) with respect to Problem (6) is obtained by utilizing the user association rule acc. best rate, i.e.,
\[ s(u, \eta) = \arg\max_{j=1, \ldots, N} c_j(u, \eta). \tag{8} \]
Note that a user associates with the BS, which provides the highest average spectral efficiency, hence occupying the fewest radio resources for data transmission. This approach is optimal from a network and energy consumption point of view, but may be suboptimal from a user point of view, in particular if the BS’s load is rather high.

2) Throughput-optimality: For \( \alpha = 1 \), we maximize the geometric mean of resources available in the network. Since acc. Problem (6), highly loaded cells and explicitly overloaded cells are avoided, users benefit from more free resources in those cells, increasing their throughput. The user association rule then becomes
\[ s(u, \eta) = \arg\max_{j=1, \ldots, N} c_j(u, \eta)(1 - \eta_j), \tag{9} \]
where the expected throughput \( c_j(u, \eta)(1 - \eta_j) \), experienced by the user and given by the product of the achievable rate and the BS’s idle time, is maximized.

3) BS switching rule: For the ESM use case with progressive BS switching, we apply the simple rule of shutting down the BS with the lowest load, obtained from the Fixed Point (5b) with the underlying user association rule according to the highest RSRP. Additionally, when switching off BSs, a minimum RSRP coverage \( C_{\min} \) has to be ensured. Hence, we choose BS \( i \) to be switched off, if
\[ i = \arg\min_j \eta_j \quad \text{s. t.} \quad C \geq C_{\min}. \tag{10} \]

4) Antenna tilt adjustment: Since, in general, the impact of varying the antenna tilts on the cell loads cannot be described analytically and in order to avoid high computational effort, we resort to the Taxi Cab Method, a coordinate-wise search algorithm already used in [12], to find a proper tilt configuration. Note that this approach may not be optimal, i.e., it may find only a local optimum with respect to Problem (6).

5) Transformation of the optimized partition: As a last step, the optimal partition \( \mathcal{P}^* \) obtained from the fixed point iteration for the optimized configuration is transformed into standard-compliant cell individual offsets (CIOs) according to the algorithm proposed in [12]. We also refer to [12] for further aspects of practical implementation, e.g., regarding measurements, information processing, and architectural aspects.
To present simulation results, we utilize the deployment, propagation, and traffic models of a real network deployed in a metropolitan area of a large North-American city.

### A. Simulation Setup and Scenario

Simulations are performed by a state-of-the-art LTE system level simulator, which implements the full user-plane functionality of LTE release 8 specifications for the PDCP, RLC, MAC, and PHY layers, and also most of the corresponding control-plane functions. The link layer performance is applied by the Mutual Information based Exponential SNR Mapping (MIESM). Table I provides the main differences between the network model used by the algorithm and the procedures implemented in the simulation tool. For the simulations, we take snapshots of the traffic intensity at every two hours and simulate at least 180 seconds of real time for each snapshot. Fig. 1 and Table II summarize the scenario, traffic characteristics, and network settings.

#### TABLE I

<table>
<thead>
<tr>
<th>Model</th>
<th>Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Traffic</strong></td>
<td>mean arrival intensity, mean file size per 15x15 m pixel</td>
</tr>
<tr>
<td><strong>User association</strong></td>
<td>optimal partition $P^*$</td>
</tr>
<tr>
<td><strong>User scheduling</strong></td>
<td>Round Robin</td>
</tr>
<tr>
<td><strong>Modulation and coding</strong></td>
<td>modified Shannon capacity $c_1(u, \eta)$, $a = 0.60$, $b = 0.45$, $c_{\text{max}} = 0.88$</td>
</tr>
<tr>
<td><strong>Link adaptation</strong></td>
<td>–</td>
</tr>
<tr>
<td><strong>Payload overhead</strong></td>
<td>considered in parameter $a$</td>
</tr>
<tr>
<td><strong>Cell loads</strong></td>
<td>calculated acc. $\eta = f(\eta)$</td>
</tr>
<tr>
<td><strong>Antenna conf.</strong></td>
<td>SISO</td>
</tr>
</tbody>
</table>

### B. Algorithms and Reference Simulations Considered

In the following, we present our simulation results according to the two ESM schemes proposed in Section III:

(A) Energy saving management throughout a day without BS switching (Section IV-C),

(B) Energy saving management with incremental BS switching acc. to Rule (10) (Section IV-D).

For each of the schemes, we optimize different network control parameters with respect to both, the energy-efficiency and the throughput-optimality objectives, i.e.,

(1) Network without optimization (denoted as INITIAL),

(2) Energy ($\alpha = 0$) / CIOs (EE CIO),

(3) Throughput ($\alpha = 1$) / CIOs (TP CIO),

(4) Energy ($\alpha = 0$) / CIOs and tilts (EE CIO+TILT),

(5) Throughput ($\alpha = 1$) / CIOs and tilts (TP CIO+TILT).

#### TABLE II

<table>
<thead>
<tr>
<th>Scenario and Network Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deployment area</strong></td>
</tr>
<tr>
<td><strong>Sites / cells</strong></td>
</tr>
<tr>
<td><strong>Inter-site distance</strong></td>
</tr>
<tr>
<td><strong>Spatial average traffic at peak hour $\sigma_{\text{peak}}$</strong></td>
</tr>
<tr>
<td><strong>Daily traffic profile</strong></td>
</tr>
<tr>
<td><strong>Spatial traffic profile</strong></td>
</tr>
<tr>
<td><strong>Mean data flow size $\Omega$</strong></td>
</tr>
<tr>
<td><strong>BS energy consumption</strong></td>
</tr>
<tr>
<td><strong>Bandwidth $B$</strong></td>
</tr>
<tr>
<td><strong>Maximum output power</strong></td>
</tr>
<tr>
<td><strong>Antenna configuration</strong></td>
</tr>
<tr>
<td><strong>Antenna patterns</strong></td>
</tr>
<tr>
<td><strong>Minimum RSRP $p_{\text{max}}$</strong></td>
</tr>
<tr>
<td><strong>Minimum RSRP coverage $C_{\text{min}}$</strong></td>
</tr>
</tbody>
</table>

### C. ESM without BS Switching

Fig. 2(a) depicts the user throughput 5 %-ile throughout a day for various optimization approaches. One can observe that the joint optimization of antenna tilts and cell individual offsets (CIO+TILT) outperforms the optimization of CIOs only (CIO), while there are only slight differences between the results for both the energy consumption (EE) and the throughput (TP) objectives. Only during the peak hour (10 pm), the optimization of CIOs with respect to the user throughput (TP CIO) is more advantageous compared to the initial network (INITIAL) and the network optimized according to minimum energy consumption (EE CIO). In that case, overload in two of the cells is avoided by shifting traffic to lower loaded neighboring cells, so that a throughput 5 %-ile of 400 kByte per second is provided. With the joint adjustment of antenna tilts and CIOs (CIO+TILT) a considerable reduction in overall network load is achieved. As a consequence thereof, the network energy consumption is lowered by up to 10 % during the peak hour in our scenario while still serving the given traffic demand, see Fig. 2(b). Since the reduction of energy consumption relies on lowering...
the network load, there is only a slight reduction in energy consumption during the low traffic times in the early morning hours. Similar to the throughput curves, the results for the joint optimization of tilts and CIOs are nearly the same for both the objectives (EE and TP), because they only differ noticeably for cell loads higher than 50%, which is rarely the case. A slight increase in energy consumption compared to the initial network configuration can be observed for the network optimized with respect to throughput by adjusting CIOs (TP CIO). Since users are forced to connect to lower loaded cells at the cost of lower spectral efficiency, this leads to a higher amount of radio resources required to serve the users’ data requests. Hence, the energy consumption increases; however, a minimum user throughput can be guaranteed.

D. ESM with BS Switching

Fig. 3(a) depicts the user throughput 5 %-ile for a well configured network at 10 pm (peak hour). For linearly decreasing BS density, the remaining BSs have to serve the additional traffic and, therefore, become more heavily loaded. In addition, the interference situation also changes from many weak to few strong interfering BSs. Hence, the 5 %-ile user throughput tends to decrease strongly. Also here, the joint optimization of antenna tilts and cell individual offsets (CIO+TILT) achieves the highest increase in user throughput, where an average gain of approx. 10% compared to the initial network performance can be observed. As expected, the optimization with respect to the user throughput (TP) yields higher 5 %-ile user throughputs in the high cell load regime, i.e., for lower base station density. This result is in accordance with the observations taken from the peak hour behavior of the algorithms, as shown in Section IV-C. Fig. 3(b) shows the average base station power consumption, which increases more than linearly with linearly decreasing BS density. This is also due to worse interference and rate conditions at the cell edges. Hence, more users at the cell edges suffer from stronger interference and, therefore, tend to require more radio resources for the same traffic demand. Concluding, the less base stations are...
active, the more resources have to be occupied, to serve the same traffic demand. This effect can slightly be mitigated by adapting the network control parameters after a BS is shut down, where, again, the joint adjustment of CIOs and tilts (CIO+TILT) leads to better results than CIOs only (CIO).

Summarizing, the joint adaptation of CIOs and antenna tilts may be highly beneficial for varying traffic demand in time and space. For very high load scenarios (> 50%), e.g., caused by aggressive switching strategies, the performance of the algorithms is limited by increasing loads of multiple neighboring cells, which makes it difficult to compensate for low user throughputs.

V. CONCLUSIONS

In this paper, we extend our framework for integrated SON, already incorporating the CCO, COC, and MLB use cases, by the energy saving management (ESM) SON use case, which may run concurrently with the other use cases. We exploit the cell load as a proxy for network performance and base station energy consumption, and estimate and predict the loads based on measurements of receive powers and traffic distributions, and with the help of a cell load model.

The ability to predict the cell loads then enables the SON to simultaneously optimize cell individual offsets and antenna downtilts towards a common objective. It can be shown that, in every case studied, a joint optimization is superior to the adjustment of CIOs only. Throughout a day, we observe remarkable improvements in user throughputs and reduction in network energy consumption for both, the user throughput and energy efficiency optimization objectives.

In addition to adapting CIOs and tilts to daily traffic variations, we also consider switching off base stations for a given traffic demand to further save energy. In general, the user throughputs then decrease rapidly due to higher loaded cells and the average BS energy consumption increases. Both effects can be mitigated by subsequently applying the same CIO and tilt adjustment algorithms. However, in very high load scenarios with multiple highly loaded neighboring cells, the performance of the algorithms is limited by less capabilities of shifting traffic.

ACKNOWLEDGMENT

The work presented in this paper was partly sponsored by the government of the Free State of Saxony, Germany, and by the European Regional Development Fund (ERDF) within the Cool Silicon Cluster of Excellence under contracts 31529/2794 and 14056/2367.

The authors would like to thank Mr. Stefan Funck and Dr. Andreas Hecker of Actix GmbH, Dresden, Germany for providing and extending the simulator to meet our needs.

REFERENCES